BUTTERWORTH HEINEMANN
0264-410X(94)00014-X
Vaccine, Vol 13, No. 6, pp. 581-591, 1995

0264-410X/95 $10.00 + 0.00

Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences

Gabriel E. Meister*, Caroline G.P. Roberts*, Jay A. Berzofsky† and Anne S. De Groot*

We have designed two computer-based algorithms for T cell epitope prediction, OptiMer and EpiMer, which incorporate current knowledge of MHC-binding motifs. OptiMer locates amphipathic segments of protein antigens with a high density of MHC-binding motifs. EpiMer identifies peptides with a high density of MHC-binding motifs alone. These algorithms exploit the striking tendency for MHC-binding motifs to cluster within short segments of each protein. Putative epitopes predicted by these algorithms contain motifs corresponding to many different MHC alleles, and may contain both class I and class II motifs, features thought to be ideal for the peptide components of synthetic subunit vaccines. In this study, we describe the use of OptiMer and EpiMer for the prediction of putative T cell epitopes from Mycobacterium tuberculosis *and human immunodeficiency virus protein antigens, and demonstrate that these two algorithms may provide sensitive and efficient means for the prediction of promiscuous T cell epitopes that may be critical to the development of vaccines against these and other pathogens.*

Keywords: MHC-binding; T-cell epitopes; vaccine

The cellular immune response to pathogens depends upon the presentation and recognition of their protein antigens in the form of intracellularly processed peptides, bound to class I or class II major histocompatibility complex (MHC) molecules and expressed at the cell surface. Peptides presented in conjunction with MHC class I molecules are derived from antigens synthesized in the cytoplasm, and are generally from 8 to 10 amino acids in length[1-3]. Peptides derived from exogenous antigens are usually presented in the context of MHC class II molecules, and range in length from around 10 to over 20 amino acids[4-8].

The identification of those peptides that stimulate T cell responses, termed T cell epitopes, is essential to the development of successful vaccines. Several methods have been employed to locate T cell epitopes within the amino acid sequences of viral and bacterial protein antigens[3,9-20]. One common approach has been to synthesize overlapping peptides which span the entire sequence of a protein antigen. These overlapping peptides are then tested for their capacity to stimulate T cell proliferative or cytotoxic responses *in vitro*[9,10]. While the overlapping peptide method is thorough, it is both cost- and labor-intensive; for a given protein of length *n* amino acids, (*n*/10)-1 peptides that are 20 amino acids long (20-mer) and overlap by 10 amino acids would need to be synthesized to employ this method of epitope identification.

Several computer-based algorithms have been designed to predict T cell epitopes from the amino acid sequences of proteins. Notably, the AMPHI algorithm searches a protein's primary structure for peptides with a high probability of folding as amphipathic structures[11,12]. In a previous analysis of the predictive power of the AMPHI algorithm, 70% of published epitopes were shown to contain sequences that would have been predicted by AMPHI[11,13]. Even as the number of known T cell epitopes has quadrupled since the advent of the AMPHI algorithm, 65% are amphipathic, and the correlation remains highly significant[12]. A new structural basis for this empirical correlation, despite the lack of helicity in peptides as bound in the MHC groove, has recently been suggested[14]. This has been based on two observations: first, that the influenza peptide bound in the groove of DR1 in the crystal structure of Stern *et al.* is a beta-strand with a 130° twist, giving it a hydrophobic periodicity similar to that of an amphipathic helix[15]; and second, that the spacing of hydrophobic anchor

*TB/HIV Research Laboratory, Brown University, Providence, RI 02912, USA. †Molecular Immunogenetics and Vaccine Research Section, Metabolism Branch, National Cancer Institute, NIH, Bethesda, MD 20892, USA.
(Received 15 August 1994; revised 4 October 1994; accepted 4 October 1994)

residues in the majority of MHC binding motifs fits the periodicity sought by AMPHI even though the peptides are not helical. Other epitope prediction algorithms which analyze protein sequences for specific secondary structural or sequence characteristics[16-19] generally search for a spacing of hydrophobic residues similar to that searched for by the AMPHI algorithm.

Another approach to T cell epitope identification has been to search protein sequences for regions that contain MHC-binding motifs[19,20], amino acid motifs found in a large proportion of peptides that bind to specific MHC alleles. Such motifs have been derived using two methods. Primarily, large pools of "naturally processed" peptides, derived from endogenously processed proteins, are acid-eluted from membrane-bound MHC molecules and sequenced. These sequences, when aligned, have been shown to incorporate certain amino acids at specific positions; these "anchor" residues are thought to facilitate peptide binding within the MHC binding groove[3]. The restricted patterns of anchor residues, termed "motifs", differ for peptides eluted from different MHC alleles, suggesting that the interaction of the anchor amino acids with the surface of the MHC molecule determines the MHC specificity of immune response to peptide epitopes.

MHC-binding motifs have also been deduced by alignment of published allele-specific T cell epitope sequences[21,22]. The literature now contains motifs for a wide variety of human class I, class II, and murine MHC alleles[3,21,44]. Certain motifs have been demonstrated to accurately predict T cell epitopes from primary structures[41-44]. However, the predictive capacities of single MHC-binding motifs vary; in some cases, peptides which bind MHC molecules lack correlation with motifs, and not all peptides containing these binding motifs are immunodominant[20,21,45-50].

Our laboratory has developed two novel algorithms, OptiMer and EpiMer, designed to predict T cell epitopes from protein primary structures. OptiMer takes into account both amphipathicity and MHC-binding motifs, while EpiMer focuses on the location of MHC-binding motifs alone.

Using published MHC-binding motifs, OptiMer examines the amino acid sequences of proteins and generates a list of peptides that contain these motifs; the algorithm then identifies peptides that would be amphipathic if folded as a helix or twisted as a beta-strand, using the AMPHI algorithm. These potentially amphipathic peptides are compared to the list of MHC-binding motif matches. OptiMer extends the predicted amphipathic peptides, to maximize the density of MHC-binding motif matches per length of protein region.

The EpiMer algorithm searches protein amino acid sequences for MHC-binding motif matches, generating a list of matches for each protein. The algorithm then identifies clusters of MHC-binding motifs, predicting putative T cell epitopes based on the relative density of these motifs. A striking observation demonstrated here is the tendency of MHC binding motifs to cluster within protein sequences, creating regions more promising for use in synthetic subunit vaccines.

These two novel algorithms, OptiMer and EpiMer, were used to predict putative epitopes in five *Mycobacterium tuberculosis* (Mtb) protein antigens (14, 16, 19, 38, and 65 kDa) and three human immunodeficiency virus (HIV) protein antigens [nef, gp160, and reverse transcriptase (RT)]. To evaluate the new algorithms' predictive power, we have compared OptiMer- and EpiMer-predicted epitopes, AMPHI-predicted epitopes, and peptides that would have been synthesized using the "overlapping peptide" method, to a selection of T cell epitopes that have been published for each of these eight proteins.

METHODS

MHC class I- and class II-specific binding motifs were collected from the literature (*Table 1*). In total, 15 distinct class II and 19 distinct class I motifs were used by the OptiMer and EpiMer algorithms. Protein primary sequences were searched for peptides that contained each MHC-binding motif using the DataMan v2.0.2 text processor (Andrew Thomas-Cramer, FeatherSoft, Madison, WI), generating a complete list of motif matches for each protein.

Eight protein sequences were obtained from the Protein Identification Resource (National Library of Medicine): A43589 (14 kDa Mtb antigen); A43823 (16 kDa Mtb antigen); SO2753 (19 kDa Mtb antigen); P15712 (38 kDa Mtb antigen); A26950 (65 kDa Mtb antigen); P04582 (HIV-1 BH8 gp160); P03406 (HIV-1 BRU nef); and P03367 (HIV-1 BRU RT). The HIV protein RT is coded for by amino acid residues 168 – 728 of the *pol* gene. MHC class II-restricted T cell epitopes for the five Mtb proteins, and MHC class I- and class II-restricted epitopes for the three HIV proteins, were compiled from the literature[51-76] (as an example, see *Figure 1a*).

Published epitopes that had been identified with the aid of the AMPHI algorithm were excluded from the list of published epitopes compiled for this study, to avoid introducing bias in favor of either AMPHI or OptiMer, both of which are based on the identification of amphipathic structures. For the three HIV proteins analyzed, only epitopes published for the strains of HIV noted above, and that had been identified in either human or murine models, were included in our analysis. (As gp160 derived from the HIV strain BH10 is 99% homologous to that of the BH8 strain, gp160 T cell epitopes identified in the BH10 strain of HIV were also included, with appropriate modifications in amino acid numbering.)

The OptiMer algorithm lengthens regions of proteins predicted by the AMPHI algorithm[11] in order to generate peptides with a maximal density of MHC-binding motif matches (*Figure 1b*). Briefly, the algorithm compares each potentially amphipathic segment, of length n amino acids, to the list of MHC-binding motif matches for the same protein. The amphipathic regions are lengthened at both their N- and C-termini until a maximal density, d, of MHC-binding motif matches ($d = m/n$, where m = the number of included motif matches) is reached, generating an OptiMer-predicted peptide (*pred*). OptiMer predictions can be customized using lists of motif matches specific to MHC class I or class II alleles, or both; likewise, these lists can include or exclude non-human MHC-binding motifs.

The OptiMer algorithm allows the user to define a level of motif density below which modified amphipathic peptides are excluded from the final list of putative epitopes. As this requisite level of motif density

is increased, the number of OptiMer-predicted peptides which would potentially be synthesized, and thus the number of amino acid residues required to construct these peptides (n_{pred}) decreases. OptiMer-predicted peptides with a motif density d greater than $\bar{d}$ were chosen for the analysis presented here, as they showed the strongest correlation to T cell epitopes published in the literature (data not shown).

The EpiMer algorithm uses the custom list of MHC-binding motif matches generated for a given protein antigen to construct a motif density "map" or histogram (*Figure 2*). By stepping a reading frame of length r one amino acid at a time through the protein primary structure, the algorithm determines the motif density d for each peptide of length r within the protein. Given a user-defined minimum density value d_{min}, itself a sum of the protein's mean MHC-binding motif density $\bar{d}$ and a positive or negative multiple of the density's standard deviation, EpiMer extracts only those motif-dense "clusters" with $d \geq d_{min}$. Finally, the algorithm uses a "threading value" t to link selected clusters into contiguous peptides, depending on their distance apart in the amino acid sequence. (As an example, $t = 5$ would assure that motif-rich clusters from 1 to 5 amino acids apart would be linked into the same predicted peptide, but that clusters 6 or more amino acids apart would not be thus linked. The technique of threading was implemented to avoid the generation of multiple peptides overlapping the same short region of a protein.) These clusters of MHC-binding motifs constitute the EpiMer algorithm's predictions for putative T cell epitopes (*Figure 1c*). EpiMer searches, like OptiMer searches, can be tailored to include or exclude non-human motifs, and to search for one class or both classes of human MHC-binding motifs.

OptiMer- and EpiMer-predicted peptides were compared to previously published epitopes for each of the eight protein antigens studied. A positive correlation was defined as an overlap of at least 11 amino acids between an OptiMer- or EpiMer-predicted peptide and a published T cell epitope, in the case of class II epitopes, and as an overlap of at least 8 amino acids, in the case of class I epitopes. These overlap values were chosen to allow the inclusion of a representative class II and class I MHC-binding motif size, respectively.

For each protein, EpiMer- and OptiMer-predicted peptides were compared to peptides derived by simulating the overlapping peptide method most

Table 1 Human class I- and class II- restricted MHC-binding motifs used by the OptiMer and EpiMer algorithms

MHC-binding motif allele	Reference	Position in peptide									
		i	*i*+1	*i*+2	*i*+3	*i*+4	*i*+5	*i*+6	*i*+7	*i*+8	*i*+9	*i*+10
HLA-A1	23			E,D	P					Y		
HLA-A2.1(a)	3		L							V		
HLA-A2.1(b)	24				A,V,I,L,Y	A,V,I,L,Y	A,V,I,L,Y					
					F,W,M,C	F,W,M,C	F,W,M,C					
HLA-A3	25		L							Y,K		
HLA-A11(a)	26		L,I,V							K		
HLA-A11(b)	3			L,I						K		
HLA-A11(c)	3				L,I					K		
HLA-A68(a)	26,27		V,S,T							R,K,H		
HLA-A68(b)	26,27		V,S,T								R,K,H	
HLA-B8(a)	3			K						L		
HLA-B8(b)	3				K					L		
HLA-B8(c)	29			R,K		R,K		W,Y		I,L		
HLA-B8(d)	23			R,K		R,K						
HLA-B27(a)	3,30		R				L,I			R,K		
HLA-B27(b)	3,30		R							R,K		
HLA-B35	31		P							Y		
HLA-B40	28		E	I,F						L		
HLA-B53(a)	31		P		E							
HLA-B53(b)	31		P			I						
HLA-DQ3.1	32	no R,K,D,E,P	no R,K,D,E	A,G,S,T	no D,E	A,V,L,I,						
HLA-DR1(a)	33	Y,F,W	no D,E	no D,E	M,L	no D,E	G,A	no D,E	no D,E	L,M,A,I,G		
										T,V,Q,S		
HLA-DR1(b)	34	Y,F			M,L		G,A			L		
HLA-DR1(c)	35	A,V,I,L,Y					S,T,A,V,I			A,V,I,L,Y		
		F,W,M,C					L,P,C			F,W,M,C		
HLA-DR(2,5,7)	33,36	Y,F,W,I L,V					S,T,A,V,G					
							I,L,P,C					R,K,H
HLA-DR2a/DR2b	37	I,L,V			D,N,Q,T							
HLA-DR3/DRw52(a)	37	F,I,L,V,Y			Q,N,R,K		R,K,H					
HLA-DR3(b)	38	A,V,I,L,Y		A,V,I,L,Y	D,E,S,T							
		F,W,M,C		F,W,M,C								
HLA-DR4(a)	37	F,L,V			M,A		T			N,Q,S,T		
HLA-DR4(b)	34	W,Y			no R,K		T,S,V,L	I,M	no R,K	no R,K,D		
HLA-DR4w4	39	V,I,L,Y								or E		
		F,W,M					N,S,T					
HLA-DR7	37	F,I,L,V,Y				H,K,R						
HLA-DR8	37	F,I,L,V,Y			M,L		R					
HLA-DRw11(5)	34	W			I,L,V			D,E				
HLA-DR17	40											

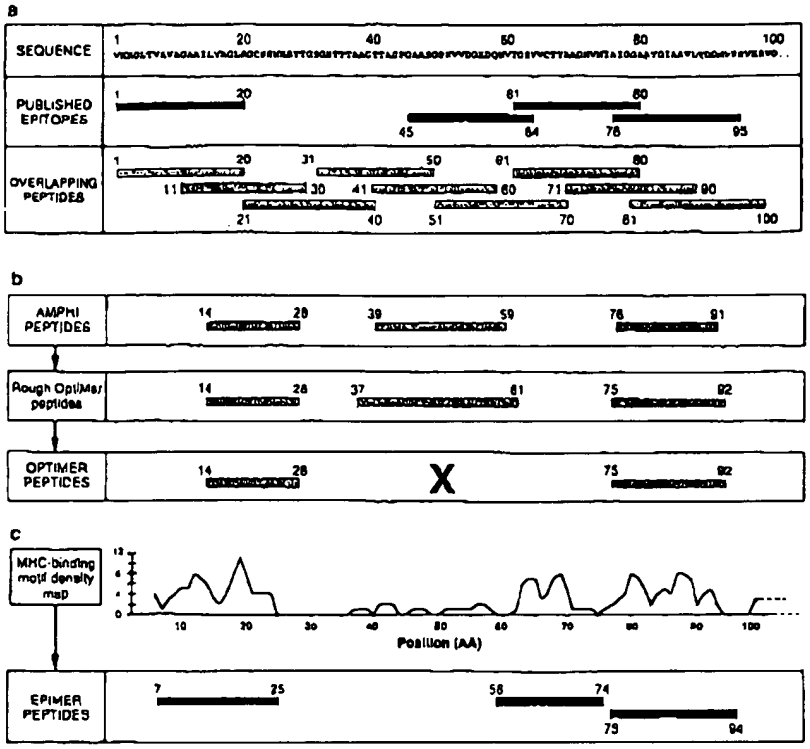

Figure 1 Schematic of methods used for locating/predicting epitopes from a model protein sequence, the 19 kDa Mtb antigen. (a) Shows the first 100 amino acids (AA) of the 19 kDa sequence, along with published epitopes and peptides that would need to be synthesized to employ the overlapping peptide method, assuming 20-mer peptides overlapping by 10 AA. For this method, 9 peptides would be needed, totaling 180 AA in length. (b) Illustrates the method for predicting OptiMer peptides. Amphipathic peptides (in this example, three AMPHI peptides for a total of 52 AA) are located, then extended to include a maximal density of MHC-binding motif matches (three Rough OptiMer peptides, totaling 58 AA); peptides with low motif densities are then systematically excluded from synthesis, generating OptiMer peptides (two peptides, totaling 33 AA). (c) Illustrates the method for predicting EpiMer peptides. Clusters of motif density are located within a protein sequence; those clusters with a high density of binding motif matches are systematically joined, dependent upon their distance apart, and selected for synthesis (giving, in this example, three peptides, totaling 56 AA). Number of binding motifs per 11-residue segment (*Y* axis) is plotted against the midpoint of an 11-residue reading frame (*X* axis), as described in Methods

commonly described in the literature[9, 10]. For this analysis, we have chosen to derive peptides of 20 residues each, overlapping by 10 residues. The AMPHI algorithm was also used to predict amphipathic peptides for our comparative analysis of prediction methods.

Three measures of the predictive power of each epitope prediction method (the overlapping peptide method, and the AMPHI, OptiMer, and EpiMer algorithms) were used in these comparisons: efficiency, E, defined as

$$E = \left(\frac{n_{corr}}{n_{prod}}\right) = \frac{\text{(total length, in amino acids, of predicted peptides correlating with published epitopes)}}{\text{(number of amino acids required to construct all predicted peptides)}} \quad (1)$$

was used to judge the capacity of each method to locate published epitopes using the fewest possible amino acid residues. Sensitivity, S, defined as

$$S = \left(\frac{pub_{corr}}{pub_{total}}\right) = \frac{\text{(number of published epitopes correlating with predicted peptides)}}{\text{(total number of published epitopes for a given protein)}} \quad (2)$$

was used to measure the ability of each method to predict a maximal number of published T cell epitopes. Finally, sensitivity per amino acid, SAA, defined as

$$SAA = \frac{S}{n_{prod}} \quad (3)$$

was used to scale sensitivity against the number of amino acids required to synthesize all peptides chosen by a specific method, giving a rough measure of cost.

For the comparisons, efficiency E, sensitivity S, and sensitivity per amino acid SAA measured for AMPHI-, OptiMer-, and EpiMer-predicted peptides, were scored against values calculated for the overlapping peptide method, to measure the relative improvement of each algorithm over this method.

The OptiMer and EpiMer algorithms were executed in Microsoft Excel v4.0 (Microsoft Corporation, Redmond, WA) using a Macintosh Quadra 650 (Apple Computer, Inc., Cupertino, CA), and are currently being implemented in the C++ programming language.

RESULTS

Prediction of putative Mtb T cell epitopes

Both the OptiMer and EpiMer algorithms were used to predict putative T cell epitopes from within the sequences of five Mtb protein antigens. (For an illustration of the methods used, see *Figures 1* and *2*.) As all published epitopes for these Mtb proteins were located through T cell proliferation assays, which measure a class II MHC-restricted response, only the list of class II-restricted MHC-binding motifs was used by OptiMer and EpiMer to predict putative epitopes.

Results for the five Mtb protein antigens studied are shown in *Table 2a*. In all, 34 T cell epitopes matching the criteria defined in Methods have been published in the literature for the five Mtb proteins studied. OptiMer generated 41 putative epitopes for these five proteins, totaling 909 amino acids in length; the EpiMer algorithm generated 42 putative epitopes, totaling 756 amino acids in length. These values are in comparison to 49 putative epitopes generated by the AMPHI

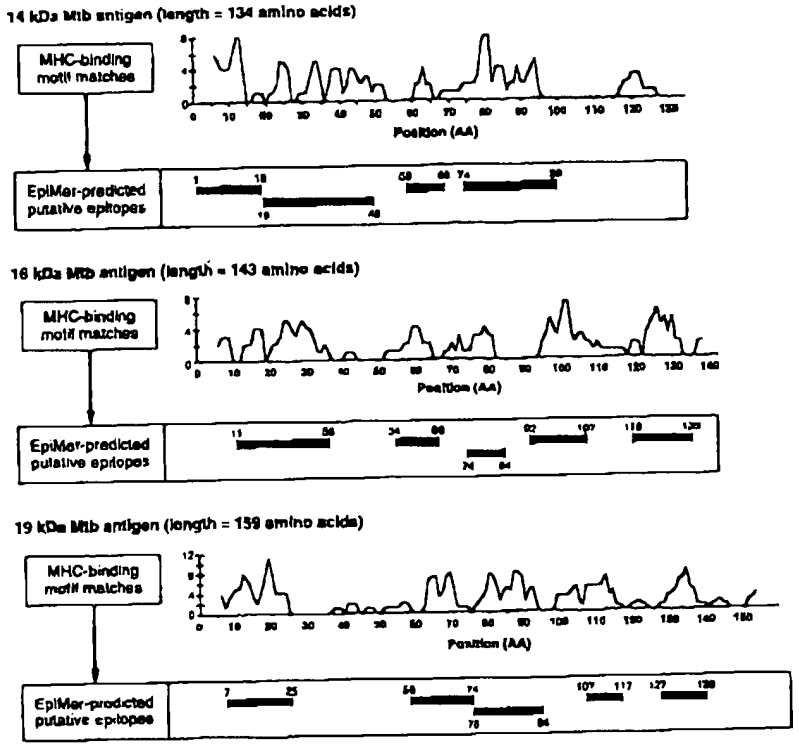

Figure 2 MHC-binding motif density "map" or histogram for three Mtb protein antigens (14, 16, 19 kDa). Number of binding motifs per 11-residue segment (*Y* axis) is plotted against the midpoint of an 11-residue reading frame (*X* axis), as described in Methods. At midpoint = 11-residue segment (*Y* axis) is plotted against the midpoint of an 11-residue reading frame (*X* axis), as described in Methods. At midpoint = residue 20 in the 19 kDa Mtb antigen, for example, 11 distinct motifs are included in the 11-residue peptide extending 5 amino acids to either side of this midpoint. For this study, a "threading value" of 10 was employed (see text): this value allows clusters of motif density ≤ 10 amino acids apart from each other to be combined into one distinct predicted putative epitope. Thus, the clusters centered at amino acids 12 and 20 of the 19 kDa antigen are linked into a EpiMer peptide extending from amino acids 7 to 25

Table 2 Correlation between AMPHI-, OptiMer- and EpiMer-predicted putative epitopes and published T cell epitopes for eight protein antigens

(a) Correlation between predicted putative epitopes and published epitopes for five Mtb protein antigens (14, 16, 19, 38, 65 kDa), using MHC class II versions of OptiMer and EpiMer

	Overlapping peptides	AMPHI peptides	OptiMer peptides	EpiMer peptides
[a]Number of peptides made	132	49	41	42
[b]Total number of amino acids required	2618	878	909	756
[c]Mean efficiency (%)	44	54	54	54
[d]Δ Efficiency over overlapping method	-	1.2-fold	1.2-fold	1.2-fold
[e]Sensitivity (%)	100 (34/34)	65 (22/34)	53 (18/34)	53 (18/34)
[f]Mean sensitivity/AA (x10[6])	2.7	5.0	4.8	5.6
[g]Δ Sensitivity/AA over overlapping method	-	1.9-fold	1.8-fold	2.1-fold

(b) Correlation between predicted putative epitopes and published epitopes for HIV protein antigens nef and gp160, using MHC class I versions of OptiMer and EpiMer

	Overlapping peptides	AMPHI peptides	OptiMer peptides	EpiMer peptides
Number of peptides made	104	36	29	30
Total number of amino acids required	2077	666	661	614
Mean efficiency (%)	35	40	41	40
Δ Efficiency over overlapping method	-	1.1-fold	1.2-fold	1.1-fold
Sensitivity (%)	100 (20/20)	80 (16/20)	75 (15/20)	50 (10/20)
Mean sensitivity/AA (x10^2)	1.6	2.9	2.7	3.1
Δ Sensitivity/AA over overlapping method	-	1.9-fold	1.7-fold	2.0-fold

(c) Correlation between predicted putative epitopes and published epitopes for HIV protein antigen RT, using MHC class I/II versions of OptiMer and EpiMer

	Overlapping peptides	AMPHI peptides	OptiMer peptides	EpiMer peptides
Number of peptides made	55	23	18	22
Total number of amino acids required	1101	433	422	361
Efficiency (%)	24	26	28	25
Δ Efficiency over overlapping method	-	1.1-fold	1.2-fold	1.1-fold
Sensitivity (%)	100 (7/7)	71 (5/7)	71 (5/7)	71 (5/7)
Sensitivity/AA (x10^2)	0.9	1.6	1.7	2.0
Δ Sensitivity/AA over overlapping method	-	1.8-fold	1.9-fold	2.2-fold

[a]Number of peptides potentially synthesized for each algorithm
[b]Total length, in amino acids, of peptides to be synthesized
[c]Mean efficiency E (as explained in Methods) for all proteins in given group
[d]Δ Efficiency = (Efficiency of prediction method) / (Efficiency of overlapping peptide method for given group of proteins)
[e]Total sensitivity, as explained in Methods, for all proteins in given group
[f]Mean sensitivity/AA (as explained in Methods) for all proteins in given group; values have been multiplied by 10^2 for the sake of clarity
[g]Δ Sensitivity/AA = (Sensitivity/AA of given method) / (Sensitivity/AA of overlapping peptide method for given group)

algorithm (totaling 878 amino acid residues), and 132 overlapping peptides (totaling over 2500 residues) needed to span each antigen using the overlapping peptide method. The OptiMer algorithm predicted published T cell epitopes for the five Mtb proteins with efficiency and sensitivity per amino acid comparable to that of the AMPHI algorithm, and both values exceeded those calculated for the overlapping peptide method. The EpiMer algorithm predicted published epitopes with an efficiency equal to that of either OptiMer or AMPHI, and again exceeding that of the overlapping peptide method; EpiMer's sensitivity per amino acid was the highest of the algorithms tested.

Prediction of putative HIV T cell epitopes

Both OptiMer and EpiMer were then used to predict T cell epitopes from within the sequences of three HIV protein antigens. Epitopes published for the HIV protein antigens nef and gp160 were almost exclusively class I MHC-restricted, while epitopes published for RT were both class I- and class II-restricted. Therefore, a version of either OptiMer or EpiMer based on the list of class I-restricted MHC-binding motifs was used to predict putative epitopes for nef and gp160, while versions of both algorithms based on the combined list of class I- and class II-restricted motifs were employed to predict putative epitopes for the HIV protein antigen RT.

Results for the HIV protein antigens nef and gp160 are shown in *Table 2b*. Twenty T cell epitopes matching the criteria set forth in Methods have been described in the literature for these two antigens. In all, 29 putative epitopes were generated by the class I-specific version of OptiMer (totaling 661 amino acids in length); 30 putative epitopes were generated by EpiMer (totaling 614 amino acids in length). AMPHI generated 36 putative epitopes (totaling 666 amino acid residues), and 104 peptides (totaling over 2000 residues in length) would have been required by the overlapping peptide method. For these two HIV protein antigens, the class I-restricted implementations of both OptiMer and EpiMer identified published epitopes with efficiency comparable to that of AMPHI, and greater than that of the overlapping peptide method. Again, the EpiMer algorithm's sensitivity per amino acid exceeded that of either the OptiMer algorithm or AMPHI.

Results for the HIV protein RT are shown in *Table 2c*. For RT, seven immunodominant T cell epitopes matching our criteria have been described in the literature. The combined class I/class II implementation of OptiMer generated 18 putative epitopes (totaling 422 amino acids); the same implementation of EpiMer generated 22 putative epitopes (totaling 361 amino acids in length). These values are in comparison to 23 putative epitopes generated by the AMPHI algorithm (totaling 433 amino acids) and 55 peptides (totaling over 1000 amino acid residues) required by the overlapping peptide method. OptiMer and EpiMer predicted published T cell epitopes for the HIV protein RT with both efficiency and sensitivity comparable to that of the AMPHI algorithm. EpiMer again achieved the highest sensitivity per amino acid of the algorithms tested.

DISCUSSION

Our laboratory has developed two algorithms, OptiMer and EpiMer, which predict putative T cell epitopes from protein primary sequences, based on secondary structural characteristics and/or the density of MHC-binding motifs within the predicted epitopes. We have compared these algorithms to previously published methods of epitope identification, and have found that the algorithms are, on the whole, able to predict T cell epitopes from protein primary structure with considerable efficiency and sensitivity per amino acid, in comparison to the overlapping peptide method (*Table 2*).

Both OptiMer and EpiMer have been designed to predict peptides which contain clusters of MHC-binding motifs (*Figure 3*). Several strongly immunodominant T cell epitopes capable of high-affinity binding to a number of different MHC molecules have been described[76, 77]; these epitopes are said to exhibit "promiscuous" or "degenerate" binding. While the OptiMer or EpiMer algorithms do not predict all possible T cell epitopes from a given protein antigen, they may preferentially predict those epitopes capable of binding to multiple MHC alleles. Thus a vaccine comprised of OptiMer or EpiMer peptides could be capable of stimulating an immune response in subjects with a variety of genetic backgrounds.

It has previously been shown that reiterative MHC-binding motifs specific to a single allele, situated within the same peptide, can greatly enhance the binding of that peptide to the associated MHC molecule[78]. The success of both OptiMer and EpiMer may be due to the striking tendency of overlapping binding motifs for multiple MHC alleles to cluster within protein antigen sequences. Indeed, as can be seen in *Figure 2*, relative densities as high as 11 motifs within a single 11-residue peptide can be located, and densities of 6 or more motifs per 11-residue peptide are not uncommon; in contrast, other regions of the same protein, as long as 35 residues, may contain very few MHC-binding motifs (*Figure 2*). These observations imply that MHC-binding motifs are not randomly distributed over a protein sequence, but rather tend to occur in clusters that may have great practical value as predictors of promiscuous MHC-binding peptides.

As described here, both OptiMer and EpiMer can be tailored to predict peptides that contain class I MHC-binding motifs, class II MHC-binding motifs, or motifs of both classes. For a pathogen that predominantly elicits class II-mediated responses, such as Mtb, class II-restricted motifs can be used to search the amino acid sequences of the pathogen's protein antigens, generating a list of peptides with a potentially broad range of activity in a variety of immunogenetic contexts. Both algorithms can also be implemented at various levels of stringency, allowing control over the density of MHC-binding motifs required to signal the location of a putative epitope, as well as over the number and mean length of putative epitopes predicted. When implemented at higher levels of stringency, OptiMer and EpiMer may be able to decrease the total cost of locating T cell epitopes within protein antigens, as well as reduce the effort required to synthesize and test these putative epitopes, in comparison to the brute force method of constructing and testing overlapping peptides.

As shown in *Table 3*, the EpiMer algorithm, based solely on the density of MHC-binding motif matches measured along a given protein sequence, has a measurably higher efficiency than either the OptiMer or AMPHI algorithm for proteins which have been extensively mapped for T cell epitopes, namely, the 19 and 65 KD antigens. This improved performance may indicate that, in fact, the EpiMer algorithm will prove to be the more useful of the two novel algorithms in the prediction of T cell epitopes. Work is underway to compare EpiMer-predicted putative epitopes to T cell epitopes published for a variety of pathogens.

As we have measured the power of the OptiMer and EpiMer algorithms by comparing their predictions to

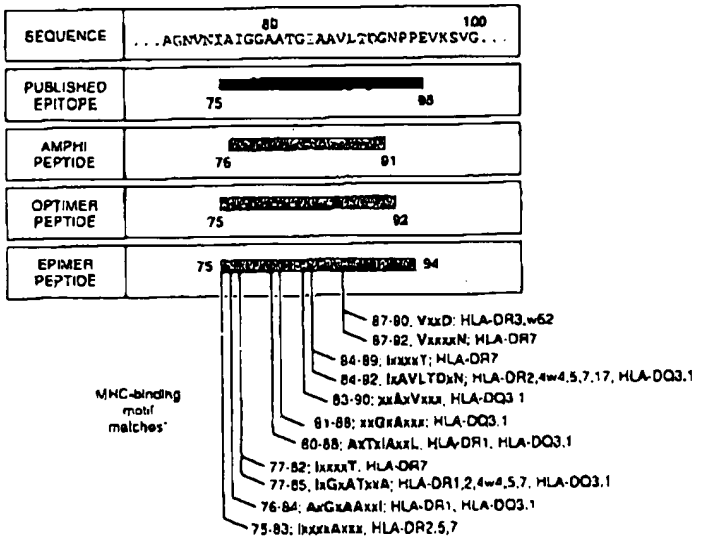

Figure 3 Comparison of published epitope, AMPHI-, OptiMer-, and EpiMer-predicted peptides, for amino acids 70–100 of the 19 kDa Mtb protein antigen sequence. MHC-binding motif matches are shown for an EpiMer-predicted putative epitope. *MHC-binding motif matches are represented as [amino acid start - amino acid stop position within protein; motif-matching amino acid sequence; alleles represented in motif-matching region, according to Table 1)

Table 3 Efficiency, sensitivity, and sensitivity per amino acid values for five individual Mtb protein antigens

Antigen	Method	Efficiency (%)	Sensitivity (%)	Sensitivity/AA*
14 kDa protein	Overlapping peptides	84	100 (7/7)	3.9
	AMPHI	100	86 (6/7)	7.8
	OptiMer	100	71 (5/7)	6.3
	EpiMer	87	57 (4/7)	6.6
16 kDa protein	Overlapping peptides	22	100 (2/2)	3.7
	AMPHI	40	50 (1/2)	6.8
	OptiMer	58	100 (2/2)	10.8
	EpiMer	45	100 (2/2)	12.0
19 kDa protein	Overlapping peptides	53	100 (8/8)	3.4
	AMPHI	57	50 (3/8)	4.9
	OptiMer	41	33 (2/8)	3.3
	EpiMer	73	50 (3/8)	6.3
38 kDa protein	Overlapping peptides	19	100 (4/4)	1.4
	AMPHI	27	100 (4/4)	3.7
	OptiMer	31	100 (4/4)	3.3
	EpiMer	12	25 (1/4)	1.2
65 kDa Protein	Overlapping peptides	43	100 (15/15)	0.8
	AMPHI	44	53 (8/15)	1.7
	Optimer	40	33 (5/15)	1.1
	EpiMer	53	53 (8/15)	1.8

Efficiency, sensitivity, and sensitivity/AA have been calculated as described in Methods
*Sensitivity/AA has been multiplied by 10^2 for clarity

published T cell epitopes, our evaluation may have been restricted by the fraction of actual epitopes that have been discovered and published to date. Predicted epitopes which do not currently appear in the literature may simply have not yet been assayed experimentally; for our analysis, such putative epitopes would have been wrongly scored as "false positive" predictions. In addition, as both AMPHI and OptiMer employ the search for amphipathic peptides, and AMPHI is often used by research groups to choose putative epitopes to test, we have excluded from our study any published T cell epitope discovered initially through the use of the AMPHI algorithm. These exclusions may have biased our evaluation against the prediction of putative epitopes by the algorithms described.

The successful prediction of putative T cell epitopes using either the OptiMer or EpiMer algorithm is dependent upon the accuracy of the MHC-binding motifs used to search the sequences in question. Not all predicted peptides can be expected either to bind to MHC molecules with high affinity, or to stimulate immune responses both *in vitro* and *in vivo*. Published motifs have been shown, in some cases, to be inaccurate predictors of either peptide–MHC-binding, immunodominance, or both; only about one-third of peptides containing the motif corresponding to a given class I MHC allele have been found to be presented by that MHC molecule[20, 29, 49, 50]. Certain motifs have been refined over time, using new information derived from amino acid substitution and peptide truncation experiments[36, 47]. While extensive experimental data confirming the accuracy of MHC-binding motifs both *in vitro* and *in vivo*, as well as data linking predicted peptide epitopes to protective immunity, are still lacking, the utility of epitope prediction for the identification of epitopes that may stimulate a protective response was recently demonstrated for a well-characterized antigen of *Plasmodium falciparum*[11].

Peptides including amino acid residues that inhibit or interfere with MHC binding have recently been described[32, 79–81]. OptiMer and EpiMer have been designed to accommodate future changes in the database of known MHC-binding motifs. As individual motifs are refined and shown to correlate with MHC-binding, immunogenicity, and the induction of protective immune responses, the usefulness of multiple motif-based epitope prediction methods such as OptiMer and EpiMer should dramatically increase. Note that the limited sensitivity of either algorithm could reflect the small number of MHC molecules for which binding motifs are known. As this number increases, the sensitivity of both algorithms may improve.

The only true test of the predictive power of the OptiMer and EpiMer algorithms will be in the synthesis and *in vitro* testing of predicted epitopes. Putative T cell epitopes predicted by either of these novel algorithms must be tested in model systems for both immunogenicity and positive correlation with immunoprotective responses. Synthesis and testing of the putative Mtb epitopes described herein are underway.

The OptiMer algorithm combines searches for secondary structural characteristics and MHC-binding motifs, and predicts amphipathic, "promiscuous" putative T cell epitopes that contain motifs for multiple MHC alleles. Notably, the EpiMer algorithm, solely based on the clustering of MHC-binding motifs within protein primary structures, also predicts putative "promiscuous" epitopes with a nearly equivalent level of predictive success. Both OptiMer and EpiMer can be tailored to search for any combination of MHC-binding motifs; the algorithms can also be modified to include new motifs as they are published. If OptiMer- or EpiMer-predicted peptides prove to function as immunodominant T cell epitopes in *in vitro* assays, as their positive correlation with published epitopes implies, these peptides may be candidates for inclusion in synthetic peptide-based vaccines. OptiMer and EpiMer may therefore provide sensitive and efficient means for the prediction of T cell epitopes crucial to the development of vaccines against Mtb, HIV, and a number of other pathogens.

NOTE ADDED IN PROOF

EpiMer predictions have been performed for a total of 20 different protein antigens; comparison with published epitopes revealed a 2.37-fold (range 0.86–3.29) improvement in sensitivity per amino acid over the overlapping method (G. Meister *et al.*, unpublished results).

Twenty-eight EpiMer-predicted Mtb peptides have been tested using PBMC from Mtb-immune donors; 16 of the 28 peptides were immunogenic *in vitro* (B. Edelson *et al.*, unpublished results).

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grant R01-AI35271 (A.S.D.). We would like to thank Paul Knopf, Brian Edelson, and Lisa Pratt for helpful reviews of the manuscript. In addition, this work could not have been completed without the support and encouragement of Glenn and Estelle Meister, and Kathie Jung Lee.

REFERENCES

1 Matsamura, M., Fremont, D.H., Peterson, P.A. and Wilson, I.A. Emerging principles for the recognition of peptide antigens by MHC class I molecules *Science* 1992, **257**, 927–934

2 Germain, R.N. and Margulies, D.H. The biochemistry and cell biology of antigen processing and presentation. *Annu. Rev. Immunol.* 1993, **11**, 403–450

3 Falk, K., Rötzschke, O., Stevanovic, S., Jung, G. and Rammensee, H.-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 1991, **351**, 290–296

4 Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A.A., Lane, W.S. and Strominger, J.L. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.* 1993, **178**, 27–47

5 Vignali, D.A.A., Urban, R.G., Chicz, R.M and Strominger, J.L. Minute quantities of a single immunodominant foreign epitope are presented as large nested sets by major histocompatibility complex class II molecules. *Eur. J. Immunol.* 1993, **23**, 1602–1607

6 Srinivasan, M., Domanico, S.Z., Kaumaya, P.T.P. and Pierce, S.K. Peptides of 23 residues or greater are required to stimulate a high affinity class II-restricted T cell response *Eur. J. Immunol* 1993, **23**, 1011–1016

7 Rudensky, A.Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. and Janeway, C.A. Sequence analysis of peptides bound to MHC class II molecules. *Nature* 1991, **353**, 622–627

8 Morrison, L.A., Lukacher, A.E., Braciale, V.L., Fan, D.P. and Braciale, T.J. Differences in antigen presentation to MHC class I- and class II-restricted influenza virus-specific cytolytic T lymphocyte clones. *J. Exp. Med.* 1986, **163**, 903–921

9 Vordermeier, H.M., Harris, D.P., Lathigra, R., Roman, E., Moreno, C. and Ivanyi, J. Recognition of peptide epitopes of the 18,000 MW antigen of *Mycobacterium tuberculosis* by murine T cells. *Immunology* 1993, **80**, 6–12

10 Ashbridge, K.R., Bäckström, B.T., Liu, H.-X., Vikerfors, T., Englebretsen, D.R., Harding, D.R. and Watson, J.D. Mapping of T helper cell epitopes by using peptides spanning the 19-kDa protein of *Mycobacterium tuberculosis*: evidence for unique and shared epitopes in the stimulation of antibody and delayed-type hypersensitivity responses. *J. Immunol.* 1992, **148**, 2248–2255

11 Margalit, H., Spouge, J.L., Cornette, J.L., Cease, K.B., DeLisi, C. and Berzofsky, J.A. Prediction of immunodominant helper T cell antigenic sites from the primary sequence *J Immunol.* 1987, **138**, 2213–2229

12 Cornette, J.L., Margalit, H., DeLisi, C. and Berzofsky, J.A. The amphipathic helix as a structural feature involved in T cell recognition. In: *The Amphipathic Helix* (Ed. Epand, R.M.). CRC Press, Boca Raton, 1993

13 Spouge, J.L., Guy, H.R., Cornette, J.L., Margalit, H., Cease, K., Berzofsky, J.A. and DeLisi, C. Strong conformational propensities enhance T cell antigenicity. *J. Immunol.* 1987, **138**, 204–212

14 Cornette, J.L., Margalit, H., Berzofsky, J.A. and DeLisi, C. Periodic variation in side chain polarities of T cell antigenic peptides: correlation with their structure and activity. (Manuscript in preparation)

15 Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L. and Wiley, D.C. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 1994, **368**, 215–221

16 Stille, C.J., Thomas, L.J., Reyes, V.E. and Humphreys, R.E. Hydrophobic strip-of-helix algorithm for selection of T cell-presented peptides. *Molec. Immunol.* 1987, **24**, 1021–1027

17 Rothbard, J.B. and Taylor, W.R. A sequence pattern common to T cell epitopes. *EMBO J.* 1988, **7**, 93–100

18 Salomon, M., Adams, S., Pardanani, A., Vazquez, S., Humphreys, R.E. and Lew, R.A. Comparison of actual and random-positioning-model distributions of peptide scavenging and T cell-presented sites in antigenic proteins. *Vaccine* 1993, **11**, 1067–1073

19 Harty, J.T. and Bevan, M.J. CD8+ T cells specific for a single nonamer epitope of Listeria monocytogenes are protective *in vivo*. *J Exp. Med.* 1992, **175**, 1531–1538

20 Pamer, E.G., Harty, J.T. and Bevan, M.J Precise prediction of a dominant class I MHC-restricted epitope of *Listeria monocytogenes*. *Nature* 1991, **353**, 852–855

21 Hobohm, U. and Meyerhans, A. A pattern search method for putative anchor residues in T cell epitopes *Eur. J. Immunol.* 1993, **23**, 1271–1276

22 Altuvia, Y., Berzofsky, J.A., Rosenfeld, R. and Margalit, H. Sequence features that correlate with MHC restriction. *Molec. Immunol.* 1994, **31**, 1–19

23 DiBrino, M., Parker, K.C., Shiloach, J., Turner, R.V., Tsuchida, T., Garfield, M., Biddison, W.E. and Coligan, J.E. Endogenous peptides with distinct amino acid anchor residue motifs bind to HLA-A1 and HLA-B8. *J. Immunol.* 1994, **152**, 620–631

24 Sette, A., Vitiello, A., Farness, P., Furze, J., Sidney, J., Claverie, J.M., Grey, H.M. and Chesnut, R. Random association between the peptide repertoire of A2.1 class I and several different DR class II molecules *J. Immunol.* 1991, **147**, 3893–3900

25 DiBrino, M., Parker, K.C., Shiloach, J., Knierman, M., Lukszo, J., Turner, R.V., Biddison, W.E. and Coligan, J.E. Endogenous peptides bound to HLA-A3 possess a specific combination of anchor residues that permit identification of potential antigenic peptides. *Proc Natl Acad. Sci. USA* 1993, **90**, 1508–1512

26 Zhang, Q.-J., Gavioli, R., Klein, G. and Masucci, M.G. An HLA-A11-specific motif in nonamer peptides derived from viral and cellular proteins. *Proc. Natl Acad. Sci. USA* 1993, **90**, 2217–2221

27 Guo, H.-C., Jardetzky, T.S., Garrett, T.P.J., Lane, W.S., Strominger, J.L. and Wiley, D.C. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature* 1992, **360**, 364–366

28 Harris, P.E., Colovai, A., Liu, Z., Dalla Favera, R. and Suciu Foca, N. Naturally processed HLA class I bound peptides from c-myc-transfected cells reveal allele-specific motifs. *J. Immunol.* 1993, **151**, 5966–5974

29 Suhrbier, A., Schmidt, C. and Fernan, A. Prediction of an HLA B8-restricted influenza epitope by motif. *Immunology* 1993, **79**, 171–173

30 Jardetzky, T.S., Lane, W.S., Robinson, R.A., Madden, D.R. and Wiley, D.C. Identification of self peptides bound to purified HLA-B27. *Nature* 1991, **353**, 326–329

31 Hill, A.V.S., Elvin, J., Willis, A.C., Aidoo, M., Allsopp, C.E., Gotch, F.M. *et al.* Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature* 1992, **360**, 434–439

32 Sidney, J., Oseroff, C., del Guercio, M.-F., Southwood, S., Krieger, J.I., Ishioka, G.Y *et al.* Definition of a DQ3.1 specific binding motif. *J. Immunol.* 1994, **152**, 4516–4525

33 Hammer, J., Takacs, B. and Sinigaglia, F. Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J. Exp. Med.* 1992, **176**, 1007–1013

34 Hammer, J., Valsasnini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B. and Sinigaglia, F. Promiscuous and allele-specific anchors in HLA-DR-binding peptides. *Cell* 1993, **74**, 197–203

35 Kropshofer, H., Max, H., Müller, C.A., Hesse, F., Stevanovic, S., Jung, G. and Kalbacher, H. Self-peptide released from class II HLA-DR1 exhibits a hydrophobic two-residue contact motif. *J. Exp. Med.* 1992, **175**, 1799–1803

36 O'Sullivan, D., Sidney, J., del Guercio, M.-F., Colón, S.M. and Sette, A. Truncation analysis of several DR binding epitopes. *J. Immunol.* 1991, **146**, 1240–1248

37 Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A.A., Lane, W.S. and Strominger, J.L. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.* 1993, **178**, 27–47

38 Sidney, J., Oseroff, C., Southwood, S., Wall, M., Ishioka, G., Koning, F. and Sette, A. DRB1*0301 molecules recognize a structural motif distinct from the one recognized by most DRB, alleles. *J. Immunol.* 1992, **149**, 2634–2640

39 Sette, A., Sidney, J., Oseroff, C., del Guercio, M.F., Southwood, S., Arrhenius, T. *et al.* HLA DR4w4-binding motifs illustrate the biochemical basis of degeneracy and specificity in peptide–DR interactions. *J. Immunol.* 1993, **151**, 3163–3170

40 Geluk, A., Van Meijgaarden, K.A., Janson, A.A.M., Drijfhout, J.W., Meloen, R.H., De Vries, R.R. and Ottenhoff, T.H. Functional analysis of DR17(DR3)-restricted mycobacterial T cell epitopes reveals DR17-binding motif and enables the design of allele-specific competitor peptides. *J. Immunol.* 1992, **149**, 2864–2871

41 Sette, A., Buus, S., Appella, E., Smith, J.A., Chesnut, R.A., Miles, C. *et al.* Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc. Natl Acad. Sci. USA* 1989, **86**, 3296–3300

42 Leighton, J., Sette, A., Sidney, J., Appella, E., Ehrhardt, C., Fuchs, S. and Adorini, L. Comparison of structural requirements for interaction of the same peptide with I-Ek and I-Ed molecules in the activation of MHC class II-restricted T cells. *J. Immunol.* 1991, **147**, 198–204

43 Corr, M., Boyd, L.F., Padlan, E.A. and Margulies, D.H. H-2Dd exploits a four residue peptide binding motif. *J. Exp. Med.* 1993, **178**, 1877–1892

44 DiBrino, M., Tsuchida, T., Turner, R.V., Parker, K.C., Coligan, J.E. and Biddison, W.E. HLA-A1 and HLA-A3 T cell epitopes derived from influenza virus proteins predicted from peptide binding motifs. *J. Immunol.* 1993, **151**, 5390–5395

45 Engelhard, V.H. Structure of peptides associated with MHC class I molecules. *Curr. Opin Immunol.* 1994, **6**, 13–23

46 Huczko, R.A., Michel, H., Sakaguchi, K., Zhu, N.Z., Shabanowitz, J., Henderson, R.A. *et al.* Characteristics of endogenous peptides eluted from class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling. *J. Immunol.* 1993, **151**, 2572–2587

47 Ruppert, J., Sidney, J., Celis, E., Kubo, R.T., Grey, H.M. and Sette, A. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* 1993, **74**, 929–937

48 Calin-Laurens, V., Trescol-Blémont, M.-C., Geriler, D. and Rabourdin-Combe, C. Can one predict antigenic peptides for MHC class I-restricted cytotoxic T lymphocytes useful for vaccination? *Vaccine* 1993, **11**, 974–978

49 Lipford, G.B., Hoffman, M., Wagner, H. and Heeg, K. Primary *in vivo* responses to ovalbumin. Probing the predictive value of the Kb binding motif. *J. Immunol* 1993, **150**, 1212–1222

50 Nijman, H.W., Houbiers, J.G.A., Vierboom, M.P.M., van der Burg, S.H., Drijfhout, J.W., D'Amaro, J. *et al.* Identification of peptide sequences that potentially trigger HLA-A2.1-restricted cytotoxic T lymphocytes *Eur. J. Immunol* 1993, **23**, 1215–1219

51 Falla, J.C., Parra, C.A., Mendoza, M., Franco, L.C., Guzman, F., Forero, J. *et al.* Identification of B- and T-cell epitopes within the MTP40 protein of *Mycobacterium tuberculosis* and their correlation with the disease course *Infect. Immun.* 1991, **59**, 2265–2273

52 Vordermeier, H.M., Harris, D.P., Lathigra, R., Roman, E., Moreno, C. and Ivanyi, J. Recognition of peptide epitopes of the 16,000 MW antigen of *Mycobacterium tuberculosis* by murine T cells. *Immunol.* 1993, **80**, 6–12

53 Lamb, J.R., Rees, A.D.M., Bal, V., Ikeda, H., Wilkinson, D., De Vries, R.R.P. and Rothbard, J.B. Prediction and identification of an HLA-DR-restricted T cell determinant in the 19-kDa protein of *Mycobacterium tuberculosis*. *Eur. J. Immunol* 1988, **18**, 973–976

54 Tan, P.L.J., Farmiloe, S., Young, J., Watson, J.D. and Skinner, M.A. Lymphocyte responses to DR4/1-restricted peptides in rheumatoid arthritis: the immunodominant T cell epitope on the 19-kd *Mycobacterium tuberculosis* protein. *Arth. Rheum.* 1992, **35**, 1419–1426

55 Rees, A.D.M., Faith, A., Roman, E., Ivanyi, J., Wiesmuller, K.-H. and Moreno, C. The effect of lipoylation on CD4 T-cell recognition of the 18,000 MW *Mycobacterium tuberculosis* antigen. *Immunology* 1993, **80**, 407–414

56 Huygen, K., Lozes, E., Gilles, B., Drowart, A., Palfliet, K., Jurion, F. *et al.* Mapping of TH1 helper T-cell epitopes on major secreted mycobacterial antigen 85A in mice infected with live *Mycobacterium bovis* BCG. *Infect. Immun.* 1994, **62**, 363–370

57 Vordermeier, H.M., Harris, D.P., Friscia, G., Roman, E., Surcel, H.-M., Moreno, C. *et al.* T cell repertoire in tuberculosis: selective anergy to an immunodominant epitope of the 38-kDa antigen in patients with active disease. *Eur. J. Immunol.* 1992, **22**, 2631–2637

58 Vordermeier, H., Harris, D.P., Roman, E., Lathigra, R., Moreno, C. and Ivanyi, J. Identification of T cell stimulatory peptides from the 38-kDa protein of *Mycobacterium tuberculosis*. *J. Immunol.* 1991, **147**, 1623–1629

59 Lamb, J.R., Ivanyi, J., Rees, A.D.M., Rothbard, J.B., Howland, K., Young, R.A. and Young, D.B. Mapping of T cell epitopes using recombinant antigens and synthetic peptides. *EMBO J.* 1987, **6**, 1245–1249

60 Anderson, D.C., van Schooten, W.C.A., Barry, M.E., Janson, A.A.M., Buchanan, T.M. and de Vries, R.A. *Mycobacterium leprae*-specific human T cell epitope cross-reactive with an HLA-DR2 peptide. *Science* 1988, **242**, 259–261

61 Haanen, J.B.A.G., Ottenhoff, T.H., Lai A Fat, R.F.M., Soebono, H., Spits, H. and de Vries, R.P. *Mycobacterium leprae*-specific T cells from a tuberculoid leprosy patient suppress HLA-DR3-restricted T cell responses to an immunodominant epitope on the 65-kDa hsp of *Mycobacteria*. *J. Immunol.* 1990, **145**, 3898–3904

62 Thole, J.E.R., van Schooten, W.C.A., Keulen, W.J., Hermans, P.W.M., Janson, A.A.M., de Vries, R. *et al.* Use of recombinant antigens expressed in *Escherichia coli* K-12 to map B-cell and T-cell epitopes on the immunodominant 65-kilodalton protein of *Mycobacterium bovis* BCG. *Inf. Immun* 1988, **56**, 1633–1640

63 Oftung, F., Mustafa, A.S., Shinnick, T.M., Houghten, R.A., Kvalheim, G., Degre, M. *et al.* Epitopes of the *Mycobacterium tuberculosis* 65-kilodalton protein antigen as recognized by human T cells. *J. Immunol.* 1988, **141**, 2749–2754

64 Fu, Y.-X., Cranfill, R., Vollmer, M., van der Zee, R., O'Brien, R.L. and Born, W. *In vivo* response of murine gamma-delta T cells to a heat shock protein-derived peptide. *Proc. Natl Acad. Sci. USA* 1993, **90**, 322–326

65 Munk, M.E., Schoel, B., Modrow, S., Karr, R.W., Young, R.A. and Kaufman, S.H.E. Lymphocytes from healthy individuals with specificity to self-epitopes shared by the mycobacterial and human 65-KD heat shock protein. *J. Immunol.* 1989, **143**, 2844–2849

66 Harris, D.P., Vordermeier, H.M., Friscia, G., Román, E., Surcel, H.-M., Pasvol, G. *et al.* Genetically permissive recognition of adjacent epitopes from the 18-kDa antigen of *Mycobacterium tuberculosis* by human and murine T cells. *J. Immunol.* 1993, **150**, 5041–5050

67 Bahraoui, E., Pagallo, M., Billaud, J.-N., Sabatier, J.-M., Guy, B., Muchmore, E. *et al.* Immunogenicity of the human immunodeficiency virus (HIV) recombinant *nef* gene product, mapping of T-cell and B-cell epitopes in immunized chimpanzees. *AIDS Res. Hum. Retro.* 1990, **6**, 1087–1098

68 Culmann, B., Gomard, E., Kieny, M.-P., Guy, B., Dreyfus, F., Saimot, A.G. *et al.* Six epitopes reacting with human cytotoxic CD8+ T cells in the central region of the HIV-1 nef protein. *J. Immunol.* 1991, **146**, 1560–1565

69 De Groot, A.S., Clerici, M., Hosmalin, A., Hughes, S.H., Barnd, D., Hendrix, C.W. *et al.* Human immunodeficiency virus reverse transcriptase T helper epitopes identified in mice and humans: correlation with a cytotoxic T cell epitope. *J. Infect. Dis.* 1991, **164**, 1058–1065

70 Hadida, F., Parrot, A., Kieny, M.-P., Sadat-Sowti, B., Mayaud, C., Debre, P. and Autran, B. Carboxyl-terminal and central regions of human immunodeficiency virus-1 nef recognized by cytotoxic T lymphocytes from lymphoid organs: an *in vitro* limiting dilution analysis. *J. Clin. Invest.* 1992, **89**, 53–60

71 Hosmalin, A., Kumar, S., Barnd, D., Houghten, R., Smith, G.E., Hughes, S.H. and Berzofsky, J.A. Immunization with soluble protein-pulsed spleen cells induces class I-restricted cytotoxic T lymphocytes that recognize immunodominant epitopic peptides from *Plasmodium falciparum* and HIV-1. *J. Immunol.* 1992, **149**, 1311–1318

72 Koenig, S., Fuerst, T.R., Wood, L.V., Woods, R.M., Suzich, J.A., Jones, G.M. *et al.* Mapping the fine specificity of a cytolytic T cell response to HIV-1 nef protein. *J. Immunol.* 1990, **145**, 127–135

73 Lieberman, J., Fabry, J.A., Kuo, M.-C., Earl, P., Moss, B. and Skolnik, P.R. Cytotoxic T lymphocytes from HIV-1 seropositive individuals recognize immunodominant epitopes in gp160 and reverse transcriptase. *J. Immunol.* 1992, **148**, 2738–2747

74 Shiral, M., Pendelton, C.D. and Berzofsky, J.A. Broad recognition of cytotoxic T cell epitopes from the HIV-1 protein with multiple class I histocompatibility molecules *J. Immunol.* 1992, **148**, 1657–1667

75 Takahashi, H., Cohen, J., Hosmalin, A., Cease, K.B., Houghten, R., Comette, J.L. *et al.* An immunodominant epitope of the human immunodeficiency virus envelope glycoprotein gp160 recognized by class I major histocompatibility complex molecule-restricted murine cytotoxic T lymphocytes. *Proc. Natl Acad. Sci. USA* 1988, **85**, 3105–3109

76 Takahashi, K., Dai, L.-C., Fuerst, T.R., Biddison, W.E., Earl, P.L., Moss, B. and Ennis, F.A. Specific lysis of human immunodeficiency virus type 1-infected cells by a HLA-A3.1-restricted CD8+ cytotoxic T-lymphocyte that recognizes a conserved peptide sequence within the gp41 subunit of the envelope protein. *Proc. Natl Acad. Sci. USA* 1991, **88**, 10277–10281

77 Panina-Brodignon, P., Tan, A., Termijtelen, A., Demotz, S., Corradin, G. and Lanzavecchia, A. Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. *Eur. J. Imunol.* 1989, **19**, 2237–2242

78 Sette, A., Sidney, J., Albertson, M., Miles, C., Colon, S.M., Pedrazzini, T. *et al.* A novel approach to the generation of high affinity class II-binding peptides. *J. Immunol.* 1990, **145**, 1809–1813

79 Klenerman, P., Rowland-Jones, S., McAdam, S., Edwards, J., Daenke, S., Lalloo, D. *et al.* Cytotoxic T cell activity antagonized by naturally occurring HIV-1 gag variants. *Nature* 1994, **369**, 403–407

80 Bertoletti, A., Sette, A., Chisari, F.V., Penna, A., Levrero, M., De Carli, M. *et al.* Natural variants of cytotoxic epitopes are T cell receptor antagonists for antiviral cytotoxic T cells. *Nature* 1994, **369**, 407–410

81 Boehncke, W.H., Takeshita, T., Pendleton, C.D., Hooughten, R.A., Sadegh Nasseri, S., Racioppi, L. *et al.* The importance of dominant negative effects of amino acid side chain substitution in peptide–MHC molecule interactions and T cell recognition *J. Immunol.* 1993, **150**, 331–341